Inferring Latent Dynamics Underlying Neural Population Activity via Neural Differential Equations

Timothy Doyeon Kim [1] **Thomas Zhihao Luo** [1] **Jonathan W. Pillow** [1 2] **Carlos D. Brody** [1 3]

Abstract

An important problem in systems neuroscience is to identify the latent dynamics underlying neural population activity. Here we address this problem by introducing a low-dimensional nonlinear model for latent neural population dynamics using neural ordinary differential equations (neural ODEs), with noisy sensory inputs and Poisson spike train outputs. We refer to this as the Poisson Latent Neural Differential Equations (PLNDE) model. We apply the PLNDE framework to a variety of synthetic datasets, and show that it accurately infers the phase portraits and fixed points of nonlinear systems augmented to produce spike train data, including the FitzHugh-Nagumo oscillator, a 3-dimensional nonlinear spiral, and a nonlinear sensory decision-making model with attractor dynamics. Our model significantly outperforms existing methods at inferring single-trial neural firing rates and the corresponding latent trajectories that generated them, especially in the regime where the spike counts and number of trials are low. We then apply our model to multi-region neural population recordings from medial frontal cortex of rats performing an auditory decision-making task. Our model provides a general, interpretable framework for investigating the neural mechanisms of decision-making and other cognitive computations through the lens of dynamical systems.

1. Introduction

The language of dynamical systems has long been used to describe neural population activity—for example, how a neural population can coordinate movement, or how it can retain working memory or integrate evidence useful for decision-making (Hopfield, 1982; Fetz, 1992; Seung, 1996; Brody et al., 2003; Machens et al., 2005; Wong & Wang, 2006; Chaudhuri & Fiete, 2016). Recent experiments showed that neural population activities across a surprisingly large number of brain regions and species can be described by a low-dimensional dynamical system (Churchland et al., 2012; Harvey et al., 2012; Mante et al., 2013; Kaufman et al., 2014; Nieh et al., 2021). This was accompanied by a growing interest in modeling neural population activity in terms of the generic dynamical system:

$$\dot{\boldsymbol{z}}(t) = \boldsymbol{f}(\boldsymbol{z}(t), \boldsymbol{u}(t), t) \tag{1}$$

where the latent variable $\boldsymbol{z} \in \mathbb{R}^L$ evolves in a low-dimensional space with dimension L. The dynamics can be perturbed by external input stimuli $\boldsymbol{u} \in \mathbb{R}^K$. Spike times of neural activities observed on a single trial can be modeled as inhomogeneous Poisson processes:

$$\begin{gathered} \boldsymbol{\lambda}(t) = \exp(\boldsymbol{C}\boldsymbol{z}(t) + \boldsymbol{d}) \\ t^{(n)}_{x,1}, t^{(n)}_{x,2}, ..., t^{(n)}_{x,\alpha(n)} \sim \text{PoissonProcess}(\lambda_n(t)) \\ n = 1, 2, ..., N \end{gathered} \tag{2}$$

where $\boldsymbol{C}$ and $\boldsymbol{d}$ are the loading matrix and bias, $t^{(n)}_{x,\cdot}$ are the spike times and $\boldsymbol{\lambda} \in \mathbb{R}^N$ are the firing rates of the observed neurons $n = 1, 2, ..., N$. $\alpha(n)$ is the total number of spike counts of the n-th neuron.

Recent studies show that a recurrent neural network (RNN) can be trained to approximate nonlinear latent dynamics (1) in a variety of brain areas, including motor and premotor cortices during reaching and other motor tasks (Pandarinath et al., 2018; Flint et al., 2020; Suresh et al., 2020). This approach marked a huge influence in thinking towards neural computation through dynamical systems (Vyas et al., 2020). However, there are a few drawbacks to this approach.

1) RNN and its dynamics are generally high-dimensional, making it difficult to interpret and gain insight into the putatively low-dimensional dynamics underlying neural activity. 2) It often requires a large number of trials to avoid overfitting and reach competitive performance against other latent

[1]Princeton Neuroscience Institute, Princeton, New Jersey [2]Department of Psychology, Princeton University, Princeton, New Jersey [3]Howard Hughs Medical Institute, Princeton University, Princeton, New Jersey. Correspondence to: Timothy Doyeon Kim <tdkim@princeton.edu>.

Proceedings of the 38^{th} International Conference on Machine Learning, PMLR 139, 2021.

variable models. 3) Unlike a typical latent variable model introduced to neuroscience that has stochasticity in the latent dynamics, the initial state of the RNN and the external input stimuli, if any, completely determine the state trajectory. While deterministic latent dynamics may be justified for neural activity in motor areas that show characteristics of autonomous dynamics, such as the supplementary motor area during cycling (Russo et al., 2020), this is not desirable in some applications, including perceptual decision-making, as the exact same input stimulus may still give rise to different choices of the animal.

Here we address these three problems by introducing a low-dimensional nonlinear model for neural population dynamics based on neural ordinary differential equations (neural ODEs; E 2017; Chen et al. 2018). First, Section 2 reviews the background and related work. Section 3 then describes our model. We show in Section 4 that: 1) our model is interpretable. By interpretable, we mean that we can obtain the phase portrait and analyze the Jacobian around the fixed points of the system. This can give useful insights into how the dynamical system does the desired computation (Sussillo & Barack, 2013; Duncker et al., 2019); 2) our model outperforms existing methods in inferring single-trial neural firing rates and the corresponding latent trajectories that generated them, even when the number of trials and the spike counts within each trial are low; 3) our model latent trajectories are not deterministic. It may take discrete noisy sensory inputs, making them suitable for applications such as decision-making. Finally, in Section 4.4, we demonstrate that our model gives interpretable phase portraits of neural dynamics when applied to the medial frontal cortex (mFC) of rats performing an auditory decision-making task.

2. Background

We review existing state-space models in neuroscience, neural ODE, and the context of our application.

2.1. Poisson Linear Dynamical System

Poisson linear dynamical system (PLDS) model approximates the latent dynamical system in (1) as a discrete-time linear dynamical system:

$$\begin{aligned} \boldsymbol{z}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_{z_0}, \boldsymbol{Q}_{z_0}) \\ \boldsymbol{z}_{k+1} &= \boldsymbol{A}\boldsymbol{z}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{\eta}_k \\ \boldsymbol{\eta}_k &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}) \end{aligned} \tag{3}$$

and uses the Poisson distribution to model spike counts that fall in each of the bins k of width Δt:

$$\begin{aligned} \boldsymbol{\lambda}_k &= \exp(\boldsymbol{C}\boldsymbol{z}_k + \boldsymbol{d}) \\ x_{n,k} &\sim \text{Poisson}(\Delta t \lambda_{n,k}) \end{aligned} \tag{4}$$

We typically apply Laplace approximation to the log-posterior of this model and optimize via Expectation-Maximization (Macke et al., 2011).

2.2. Latent Factor Analysis via Dynamical Systems

Latent factor analysis via dynamical systems (LFADS; Sussillo et al. 2016; Pandarinath et al. 2018) approximates (1) with an RNN, where $\boldsymbol{h} \in \mathbb{R}^H$ is the hidden state of the gated recurrent unit (GRU; Cho et al. 2014) of dimension H:

$$\begin{aligned} \boldsymbol{h}_0 &\sim \mathcal{N}(\boldsymbol{\mu}_{h_0}, \boldsymbol{Q}_{h_0}) \\ \boldsymbol{h}_{k+1} &= \text{GRU}(\boldsymbol{h}_k, \boldsymbol{u}_k) \\ \boldsymbol{z}_k &= \boldsymbol{W}\boldsymbol{h}_k + \boldsymbol{b} \end{aligned} \tag{5}$$

(5) is called the generator. The low-dimensional latent factors $\boldsymbol{z}$ are mapped to neural observations in the same way as (4). LFADS also has a black box recognition model called the encoder that serves to amortize inference, where it gets a sequence of spike counts $\boldsymbol{x}_{1:K}$ as an input and outputs parameters $\boldsymbol{\mu}_{h_0}$ and $\boldsymbol{Q}_{h_0}$. Typically, the encoder is a bidirectional GRU that runs forward and backward in time:

$$[\boldsymbol{\mu}_{h_0}, \boldsymbol{Q}_{h_0}] = \text{Encoder}(\boldsymbol{x}_{1:K}) \tag{6}$$

When $\boldsymbol{u}$ is not directly observed, we may replace it with outputs from an extra RNN called the controller that gets outputs from the encoder as its input. Recently, AutoLFADS, which optimizes hyperparameters of LFADS via population-based training, has been developed to increase the performance of LFADS (Keshtkaran et al., 2021).

2.3. Neural Ordinary Differential Equations

An alternative to approximating (1) with an RNN is to approximate $\boldsymbol{f}$ with a feedforward neural network (FNN) and use an ODE solver to obtain the latent trajectory. Broadly, there are two ways to obtain the gradient of the loss with respect to the parameters ψ of the FNN. First, we can perform reverse-mode differentiation through the solver (called discrete adjoint sensitivity). Some find that, under certain conditions, this may be more performant than the second approach, continuous adjoint sensitivity (Rackauckas et al., 2018). Continuous adjoint sensitivity computes the gradient by realizing that the gradient of the loss follows another ODE (Pontryagin et al., 1962), and solving this ODE backwards in time (Serban & Hindmarsh, 2005; Chen et al., 2018). If we define the adjoint states

$$\boldsymbol{a}(t) = \frac{d\mathcal{L}}{d\boldsymbol{z}(t)}, \qquad \boldsymbol{a}_\psi(t) = \frac{d\mathcal{L}}{d\psi} \tag{7}$$

then they follow the ODEs

$$\begin{aligned} \frac{d\boldsymbol{a}(t)}{dt} &= -\boldsymbol{a}(t)\frac{\partial \boldsymbol{f}_\psi(\boldsymbol{z}(t), t)}{\partial \boldsymbol{z}(t)} \\ \frac{d\boldsymbol{a}_\psi(t)}{dt} &= -\boldsymbol{a}(t)\frac{\partial \boldsymbol{f}_\psi(\boldsymbol{z}(t), t)}{\partial \psi} \end{aligned} \tag{8}$$

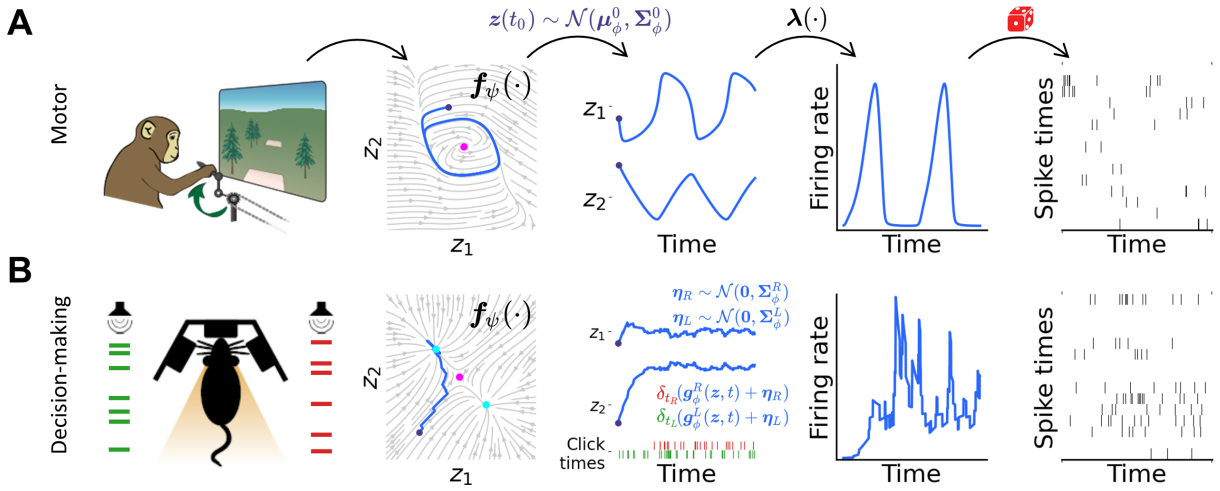

Figure 1. Generative processes of PLNDE in different scenarios. (A) An animal may perform motor tasks such as reaching or cycling that do not have explicit external input stimulus. The phase portrait of the dynamical system is inferred by $\boldsymbol{f}_\theta$. Each trial of the task is a single trajectory in this phase portrait, which starts from an initial value $\boldsymbol{z}(t_0)$, indicated by the dark blue circle. The magenta circle indicates an unstable fixed point of the dynamics. Each trajectory is mapped to the rates of Poisson processes that generate spike times of the neurons observed during the task. Monkey illustration adapted from Russo et al. 2020. (B) Neural dynamics may be affected by external sensory stimuli in tasks such as perceptual decision-making. The dynamics of the animal doing such tasks may involve two stable fixed points (cyan circles), and one unstable fixed point (magenta circle). When inputs are discrete, these inputs may affect the latent trajectories as discrete jumps at the input times. Rat illustration adapted from Brunton et al. 2013.

In Chen et al. 2018, a continuous adjoint sensitivity method that takes $\mathcal{O}(1)$ memory is discussed. This is possible via augmenting the ODE we want to solve with the ODEs in (8) and integrating the augmented ODE backwards. The backward solution of the ODE we want to solve can diverge from the forward solution with this approach if the solver is not time-reversible and uses adaptive time steps. However, time-reversibility of the solver may not be necessary for convergence under certain conditions.

Latent neural ODE models have been explored in Chen et al. 2018 and Rubanova et al. 2019 where they were applied to a wide variety of datasets with irregularly-sampled observations, for example, medical results from patients in ICU.

2.4. Handling Discrete Jumps in the Continuous Adjoint Sensitivity Method

Suppose that at a discrete event time $t_{u,j}$, our latent state is perturbed by the event such that:

$$\boldsymbol{z}\left(t_{u,j}^+\right) = \boldsymbol{z}\left(t_{u,j}\right) + \boldsymbol{g}_\phi^u\left(\boldsymbol{z}\left(t_{u,j}\right), t_{u,j}\right) \tag{9}$$

where $\boldsymbol{g}_\phi^u$ is some function of parameters ϕ that determines the update in our state $\boldsymbol{z}$. Then, the adjoint states satisfy:

$$\boldsymbol{a}\left(t_{u,j}\right) = \boldsymbol{a}\left(t_{u,j}^+\right) + \boldsymbol{a}\left(t_{u,j}^+\right)\frac{\partial \boldsymbol{g}_\phi^u\left(\boldsymbol{z}\left(t_{u,j}\right), t_{u,j}\right)}{\partial \boldsymbol{z}(t_{u,j})}$$

$$\boldsymbol{a}_\Theta\left(t_{u,j}\right) = \boldsymbol{a}_\Theta\left(t_{u,j}^+\right) + \boldsymbol{a}\left(t_{u,j}^+\right)\frac{\partial \boldsymbol{g}_\phi^u\left(\boldsymbol{z}\left(t_{u,j}\right), t_{u,j}\right)}{\partial \Theta} \tag{10}$$

where $\{\psi, \phi\} \subset \Theta$. (10) are due to Equation (13) from Jia & Benson 2019 and Remark 2 from Corner et al. 2018.

2.5. Pulsatile Evidence Accumulation in Perceptual Decision-Making

When perceptual evidence is discrete, one can discern different kinds of noise that may be at play during decision-making (e.g., noise in the latent variable versus noise in the perceived evidence) (Brunton et al., 2013). To achieve this, an experimental setup was devised where an animal (typically a rodent) is presented with randomly-timed auditory pulses to its left and right sides and is rewarded for correctly reporting which side had the greater total number of "clicks". This paradigm is often used to study perceptual decision-making (Scott et al., 2015; Pinto et al., 2018; Nieh et al., 2021).

Decision dynamics in this paradigm can be modeled as an Ornstein–Uhlenbeck process with jumps (Brunton et al., 2013; Piet et al., 2018; DePasquale et al., 2019):

$$\begin{aligned} dz = \lambda z dt &+ \delta_{t_R}\left(C^R + \eta_R\right) dt \\ &+ \delta_{t_L}\left(C^L + \eta_L\right) dt + \sigma_z dW \end{aligned} \tag{11}$$

Here, $\eta_R, \eta_L \sim \mathcal{N}(0, \sigma_s)$. C^R and C^L represent how much

on average the latent variable z should be updated at each click event coming from the right and left sides, respectively.

3. Poisson Latent Neural Differential Equations for Spiking Neural Data

We develop Poisson Latent Neural Differential Equations (PLNDE), which infers the underlying nonlinear dynamics of neural population spike trains. This framework can be applied to scenarios where there are no external inputs to the dynamics and dynamics are autonomous (Figure 1A).

Our model can also be applied to scenarios where there are external inputs to the dynamics and dynamics are non-autonomous. While our model can take sensory input stimuli $\boldsymbol{u}$ that are either continuous or discrete, our model can take sensory noise into account only when the stimuli are discrete, as in Section 2.5 and Figure 1B, or when the continuous stimuli permit noise only at finite time points.

The marginal log-likelihood, or evidence, $\log p_\Theta(\boldsymbol{t}_x|\boldsymbol{u})$ can be maximized by computing the variational lower bound $\mathcal{L}$ of $\log p_\Theta(\boldsymbol{t}_x|\boldsymbol{u})$, where $\boldsymbol{t}_x$ are the spike times of N observed neurons, $\boldsymbol{u}$ is the input, and the parameters of our model are $\Theta = \{\theta, \phi\}$. If spike times are generated as in (2), $\theta = \{\psi, \boldsymbol{C}, \boldsymbol{d}\}$ are the generative parameters and ϕ are the variational parameters.

$$\mathcal{L} = \mathbb{E}_q[\log p_\theta(\boldsymbol{t}_x|\boldsymbol{z}, \boldsymbol{u})] - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{t}_x, \boldsymbol{u})||p(\boldsymbol{z}|\boldsymbol{u})) \tag{12}$$

Let the discrete input event times be $\boldsymbol{t}_u$. In such a case, we may model (1) as:

$$\begin{aligned} \dot{\boldsymbol{z}} &= \boldsymbol{f}_\psi(\boldsymbol{z}, t) + \delta_{t_u}(\boldsymbol{g}_\phi^u + \boldsymbol{\eta}_u) \\ \boldsymbol{\eta}_u &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_\phi^u) \end{aligned} \tag{13}$$

where $\boldsymbol{f}_\psi$ is a generic function that may be approximated with a feedforward neural network (FNN). $\boldsymbol{f}_\psi$ characterizes the phase portrait, with its zeros representing the fixed points. $\boldsymbol{g}_\phi^u$ is the mean update in the latent state at each event time $t_{u,j}$, and the diagonal matrix $\boldsymbol{\Sigma}_\phi^u$ represents sensory noise in the update. $\boldsymbol{\eta}_u$ models noise in the sensory input perceived by the animal, similar to Brunton et al. 2013.

If the mapping from $\boldsymbol{z}(t)$ to $\boldsymbol{\lambda}(t)$ is defined by (2),

$$\begin{aligned} \mathbb{E}_q[\log p_\theta(\boldsymbol{t}_x|\boldsymbol{z}, \boldsymbol{u})] &= \mathbb{E}_q[\log p_\theta(\boldsymbol{t}_x|\boldsymbol{\lambda})] \\ &= \sum_{n=1}^{N} \mathbb{E}_q[\log p_\theta(\boldsymbol{t}_x^{(n)}|\lambda_n)] \end{aligned}$$

where

$$\log p(\boldsymbol{t}_x^{(n)}|\lambda_n) = -\int_{\mathcal{T}} \lambda_n(t)dt + \sum_{i=1}^{\alpha(n)} \log \lambda_n(t_{x,i}^{(n)}) \tag{14}$$

We Monte Carlo estimate $\mathbb{E}_q[\log p_\theta(\boldsymbol{t}_x^{(n)}|\lambda_n)]$. While it is possible to compute (14) via an ODE solver (Rubanova et al., 2019), we find that approximating (14) via (4) works well in practice—binning the spike times into k bins of width Δt and inferring the mean $\boldsymbol{\lambda}_k$ of the Poisson distribution works well in the synthetic datasets we test, and may be more efficient for our application as the number of neurons N can get large.

We find parameters Θ that maximize $\mathcal{L}$ via a first-order method such as ADAM.

3.1. Computing the Adjoints with Stochastic Jumps

We compute $\partial\mathcal{L}/\partial\Theta$ by using the reparametrization trick, storing $\epsilon \sim \mathcal{N}(0, I)$ instantiated in the forward pass, and using the same noise instances during the backward pass. A similar approach has been used in fitting circuit models in neuroscience (Duan et al., 2021).

$$\boldsymbol{z}\left(t_{u,j}^+\right) = \boldsymbol{z}\left(t_{u,j}\right) + \boldsymbol{g}_\phi^u + \epsilon \odot \sqrt{\boldsymbol{\Sigma}_\phi^u} \tag{15}$$

The adjoints for (15) then are

$$\begin{aligned} \boldsymbol{a}\left(t_{u,j}\right) = \boldsymbol{a}\left(t_{u,j}^+\right) &+ \boldsymbol{a}\left(t_{u,j}^+\right) \frac{\partial \boldsymbol{g}_\phi^u}{\partial \boldsymbol{z}\left(t_{u,j}\right)} \\ &+ \epsilon \odot \boldsymbol{a}\left(t_{u,j}^+\right) \frac{\partial \sqrt{\boldsymbol{\Sigma}_\phi^u}}{\partial \boldsymbol{z}\left(t_{u,j}\right)} \end{aligned}$$

$$\begin{aligned} \boldsymbol{a}_\Theta\left(t_{u,j}\right) = \boldsymbol{a}_\Theta\left(t_{u,j}^+\right) &+ \boldsymbol{a}\left(t_{u,j}^+\right) \frac{\partial \boldsymbol{g}_\phi^u}{\partial \Theta} \\ &+ \epsilon \odot \boldsymbol{a}\left(t_{u,j}^+\right) \frac{\partial \sqrt{\boldsymbol{\Sigma}_\phi^u}}{\partial \Theta} \end{aligned} \tag{16}$$

Computing the adjoints take $\mathcal{O}(M)$ memory and $\mathcal{O}(\tilde{M})$ time, similar to continuous adjoint sensitivity methods that utilize checkpointing (Serban & Hindmarsh, 2005; Rackauckas et al., 2020). M is the number of jump events and $\tilde{M}$ is the number of function evaluations.

3.2. Computing the Kullback-Leibler Divergence

The posterior $q_\phi(\boldsymbol{z}|\boldsymbol{t}_x, \boldsymbol{t}_u)$ is given by

$$\begin{aligned} q_\phi(\boldsymbol{z}|\boldsymbol{t}_x, \boldsymbol{t}_u) &= q_\phi(\boldsymbol{z}(t_0), \boldsymbol{z}(t_{u,1}), ..., \boldsymbol{z}(t_{u,M})|\boldsymbol{t}_x) \\ &= q_\phi(\boldsymbol{z}(t_0)|\boldsymbol{t}_x) \prod_{j=1}^{M} q_\phi(\boldsymbol{z}(t_{u,j})|\boldsymbol{z}(t_{u,j-1}), \boldsymbol{t}_x) \end{aligned} \tag{17}$$

$$q_\phi(\boldsymbol{z}(t_0)|\boldsymbol{t}_x) = \mathcal{N}(\boldsymbol{\mu}_\phi^0, \boldsymbol{\Sigma}_\phi^0) \tag{18}$$

$$q_\phi(\boldsymbol{z}(t_{u,j})|\boldsymbol{z}(t_{u,j-1}), \boldsymbol{t}_x) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_\phi^u) \tag{19}$$

$$\boldsymbol{\mu} = \boldsymbol{g}_\phi^u + \boldsymbol{z}(t_{u,j-1}) + \int_{t_{u,j-1}}^{t_{u,j}} f_\psi(\boldsymbol{z}(t), t)dt \tag{20}$$

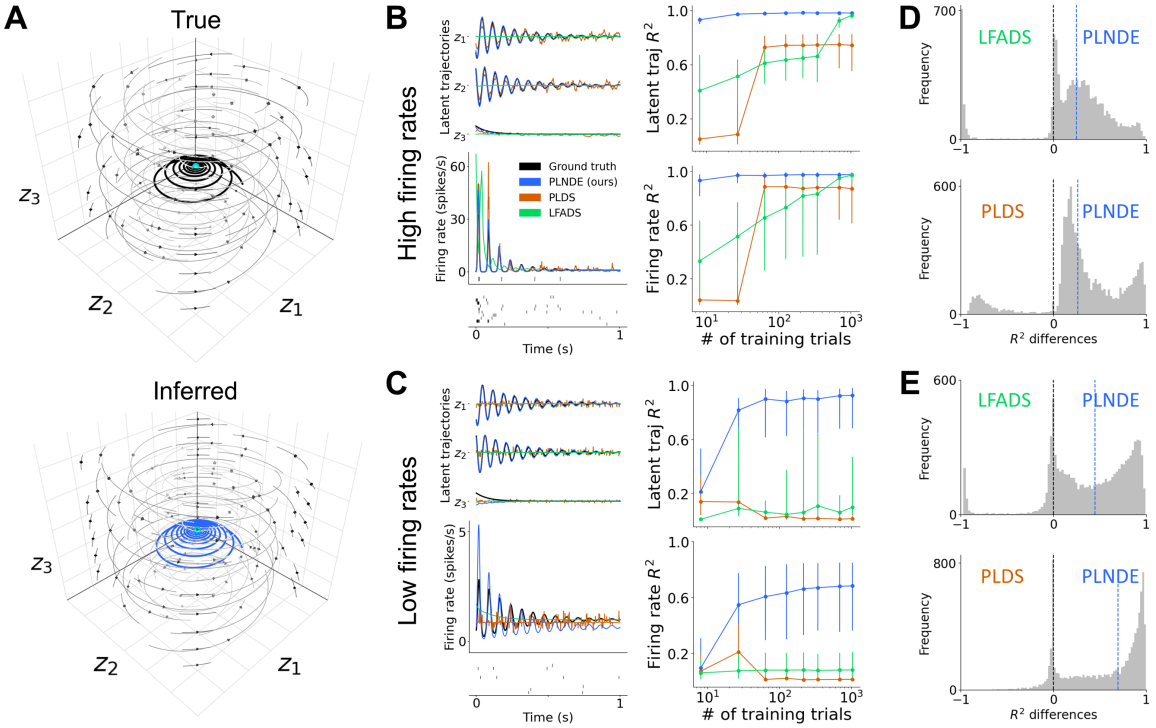

Figure 2. PLNDE accurately infers phase portrait and fixed point. PLNDE outperforms LFADS and PLDS in inferring single-trial latent trajectories and individual neural firing rates for nonlinear spiral dynamics. (A) Top: True phase portrait of the spiral. Black line indicates a trajectory in an example test trial. Bottom: Inferred phase portrait of the spiral affine-transformed to match the true phase portrait. Blue line indicates the inferred trajectory. Cyan circles indicate the true and inferred stable fixed points. (B) Condition where the mean population firing rate is high. The example test trajectories in A are shown unrolled in time, along with the corresponding firing rate of an example neuron. The first 10 example neurons' spike times are also shown. Each circle indicates the median R^2 of the latent trajectories. The error bar indicates the first and third quartiles. (C) Same as B for low mean population firing rates. (D) Summary for high firing rates, including all numbers of trials in B, and computing PLNDE R^2 − LFADS/PLDS R^2 (abscissa). The vertical dashed lines represent the medians. (E) Same as D for low firing rates.

where M is the total number of jump events, t_0 is the start of the trial, and $t_{u,j}$ is the timing of the j-th event. (18) determines the distribution of initial values. Similarly, the prior is given by

$$\begin{aligned} p(\boldsymbol{z}|\boldsymbol{t}_u) &= p(\boldsymbol{z}(t_0), \boldsymbol{z}(t_{u,1}), ..., \boldsymbol{z}(t_{u,M})) \\ &= p(\boldsymbol{z}(t_0)) \prod_{j=1}^{M} p(\boldsymbol{z}(t_{u,j})|\boldsymbol{z}(t_{u,j-1})) \end{aligned} \tag{21}$$

$$p(\boldsymbol{z}(t_0)) = \mathcal{N}(\boldsymbol{\mu}^0_{\text{prior}}, \boldsymbol{\Sigma}^0_{\text{prior}}) \tag{22}$$

$$p(\boldsymbol{z}(t_{u,j})|\boldsymbol{z}(t_{u,j-1})) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}^u_{\text{prior}}) \tag{23}$$

$$\boldsymbol{\mu} = \boldsymbol{g}^u_{\text{prior}} + \boldsymbol{z}(t_{u,j-1}) + \int_{t_{u,j-1}}^{t_{u,j}} f_\psi(\boldsymbol{z}(t), t)dt \tag{24}$$

where $\boldsymbol{\mu}$ in Equation (24) is similar to (20), with $\boldsymbol{g}^u_\phi$ replaced by $\boldsymbol{g}^u_{\text{prior}}$.

The Kullback-Leibler (KL) divergence then becomes

$$\begin{aligned} &D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{t}_x, \boldsymbol{t}_u)||p(\boldsymbol{z}|\boldsymbol{t}_u)) = \frac{1}{2}\Big[\log\frac{|\boldsymbol{\Sigma}^0_{\text{prior}}|}{|\boldsymbol{\Sigma}^0_\phi|} - d \\ &+ \text{tr}\{\boldsymbol{\Sigma}^{0^{-1}}_{\text{prior}}\boldsymbol{\Sigma}^0_\phi\} + (\boldsymbol{\mu}^0_\phi - \boldsymbol{\mu}^0_{\text{prior}})^\top \boldsymbol{\Sigma}^{0^{-1}}_{\text{prior}}(\boldsymbol{\mu}^0_\phi - \boldsymbol{\mu}^0_{\text{prior}})\Big] \\ &+ \sum_{j=1}^{M} \frac{1}{2}\Big[\log\frac{|\boldsymbol{\Sigma}^u_{\text{prior}}|}{|\boldsymbol{\Sigma}^u_\phi|} - d + \text{tr}\{\boldsymbol{\Sigma}^{u^{-1}}_{\text{prior}}\boldsymbol{\Sigma}^u_\phi\} \\ &+ (\boldsymbol{g}^u_\phi - \boldsymbol{g}^u_{\text{prior}})^\top \boldsymbol{\Sigma}^{u^{-1}}_{\text{prior}}(\boldsymbol{g}^u_\phi - \boldsymbol{g}^u_{\text{prior}})\Big] \end{aligned} \tag{25}$$

When $M = 0$, we have a model for scenarios as in Figure 1A, and when $M > 0$, scenarios as in Figure 1B. See Section A of Supplementary Material for detailed derivations.

4. Experiments

We fit PLNDE (ours), PLDS[1] (Section 2.1) and LFADS[2] (Section 2.2) to synthetic datasets with three different nonlinear dynamical systems. We generated each latent trajectory (i.e., a trial) from a different initial value of the dynamical systems. We split the data into training and test, where test trajectories were generated from initial values not seen in training. For LFADS and PLDS, we used the default initializations, hyperparameters and termination criteria given in their original repositories. We initialized the parameters of PLNDE with Glorot normal and trained for ≈ 5000 iterations. See Section B of Supplementary Material for details on how the data were generated, and the specifics of model architectures, initializations and hyperparameters. We also fit PLNDE to multiple areas within the mFC of a rat performing the decision-making task in Section 2.5 (Figure 1B; Brunton et al. 2013).

4.1. Spiral Dynamics

Figure 2 shows the results of fitting PLNDE, PLDS and LFADS to a nonlinear dynamical system that generates spirals similar to the example in Brunton et al. 2016.

Specifically, spike times of 150 neurons were generated from a system following the equations:

$$\begin{aligned}\dot{z}_1 &= -4z_1^3 - 4z_1 - 80z_2^3 - 80z_2 \\ \dot{z}_2 &= 80z_1^3 + 80z_1 - 4z_2^3 - 4z_2 \\ \dot{z}_3 &= -12z_3^3 - 12z_3\end{aligned} \tag{26}$$

The nonlinear spiral has one stable fixed point at the origin (see Figure 2A for phase portrait and fixed point). While we cannot plot the inferred nonlinear phase portrait with LFADS or PLDS, PLNDE can plot the phase portrait, providing an interpretable representation of the latent trajectories. We found that PLNDE accurately infers the phase portrait and the stable fixed point of the spiral with a sufficient number of trials (Figure 2A–C). The stable fixed point was found by finding the zeros of our FNN $\boldsymbol{f}_\psi$ with Newton's method and using automatic differentiation to find the Jacobian around the fixed point. The computed eigenvalues of the Jacobian around the fixed point inferred by PLNDE in one example session were $[-8.14, -2.91 - 79.33i, -2.91 + 79.33i]$, while the true eigenvalues were $[-12, -4 - 80i, -4 + 80i]$.

We considered how the mean population firing rates and the number of training trials affect the models' performance in recovering the true single-trial firing rates and the latent trajectories (Figure 2B–C). R^2 between the ground truth latent trajectory and the inferred latent trajectory were computed separately for each trial and each latent dimension. When the mean population firing rate is high (≈ 6.62 spikes/s), PLNDE was able to reconstruct the individual test latent trajectories with the median R^2 of 0.93 with as low as 8 training trials, while for LFADS and PLDS, median R^2 were 0.41 and 0.05, respectively (Figure 2B). As the number of training trials increased, the median R^2 increased for all models. PLNDE outperformed LFADS and PLDS in all of the different numbers of training trials that we considered. The median R^2 of PLNDE was significantly higher than the median R^2 of LFADS and PLDS with $p < 0.0001$ for all numbers of training trials (two-sided Mann-Whitney U tests; Figure 2B, D). We found similar trends as a function of the number of training trials when we computed each neuron's firing rate reconstruction accuracy (Figure 2B). R^2 was computed between the ground truth firing rate and the inferred firing rate for each trial and neuron. These results were more striking when the mean population firing rate was low (≈ 1.12 spikes/s). Even when LFADS and PLDS could not recover the latent trajectories and infer the firing rates of the neurons, with a sufficient number of trials, PLNDE was able to recover the latent trajectories and firing rates (Figure 2C). Generally, PLNDE outperformed LFADS and PLDS when the mean population firing rates were both high and low, and across different numbers of training trials (Figure 2D–E).

In these experiments, we did not optimize or train the latent dimensionality; we chose the latent dimensionality of all models we tested to be equal to the ground truth. However, we often do not know the true latent dimensionality of the system we observe. In such a case, the optimal dimensionality may be recovered by comparing the test loss (ELBO) and the test log-likelihood (LL) of each neuron across models that assume different latent dimensions. We demonstrate that recovering the optimal dimensionality with this approach is possible for the nonlinear spiral dataset in Figure 2 (Section B of Supplementary Material).

4.2. FitzHugh-Nagumo

Spike trains of 50 neurons were generated from a system governed by the FitzHugh-Nagumo dynamics:

$$\begin{aligned}\dot{z}_1 &= \rho\tau\left(z_1 - \frac{1}{3}z_1^3 - z_2\right) + I_{\text{input}} \\ \dot{z}_2 &= \frac{\tau}{\rho}\left(z_1 + a - bz_2\right)\end{aligned} \tag{27}$$

where $a = b = 0$, $\rho = 2$ and $\tau = 15$. We found that PLNDE can accurately infer the phase portrait of FitzHugh-Nagumo, capturing the limit cycle and the unstable fixed point. We trained PLNDE, PLDS and LFADS on different numbers of training trials and found that, similar to Figure 2, PLNDE outperformed LFADS and PLDS in all numbers of train-

[1]Code from: https://bitbucket.org/mackelab/pop_spike_dyn/src/master/

[2]Code from: https://github.com/tensorflow/models/tree/master/research/lfads

ing trials we considered. See Section B of Supplementary Material for details.

4.3. Mutual Inhibition with External Input Stimuli

Spike trains of 150 neurons were generated from a system with the following attractor dynamics (Wong & Wang, 2006; Piet et al., 2017):

$$\begin{aligned}\dot{z}_1 &= \tau\left(-z_1 + \frac{1}{1+\exp\left(k\left(z_2-\gamma\right)\right)}\right) \\ &\quad + \left(\delta_{t_R}\left(C_1^R + \eta_{R,1}\right) + \delta_{t_L}\left(C_1^L + \eta_{L,1}\right)\right) \\ \dot{z}_2 &= \tau\left(-z_2 + \frac{1}{1+\exp\left(k\left(z_1-\gamma\right)\right)}\right) \\ &\quad + \left(\delta_{t_R}\left(C_2^R + \eta_{R,2}\right) + \delta_{t_L}\left(C_2^L + \eta_{L,2}\right)\right)\end{aligned} \tag{28}$$

where $\tau = 10$, $k = 16$, and $\gamma = 0.5$. 30 Hz input pulses of dimension $K = 2$ affect the system. A pulse in the first dimension induces a jump in the latent dynamics with the mean being $\boldsymbol{C}^R$, while a pulse in the second dimension induces a jump with the mean being $\boldsymbol{C}^L$, where

$$\boldsymbol{C}^R = \begin{bmatrix} 0.05 \\ -0.05 \end{bmatrix}, \quad \boldsymbol{C}^L = \begin{bmatrix} -0.05 \\ 0.05 \end{bmatrix}. \tag{29}$$

Each pulse had sensory noise, where $\eta_L, \eta_R \sim \mathcal{N}(\mathbf{0}, \kappa \boldsymbol{I})$ with $\kappa = 0.001$.

We found that, similar to the spiral (Section 4.1; Figure 2) and FitzHugh-Nagumo (Section 4.2), PLNDE can infer the phase portrait of the mutual inhibition dynamics, correctly identifying the two stable and one unstable fixed points (Figure 3A). We found that PLNDE outperformed LFADS and PLDS in inferring neural firing rates and latent trajectories (Section B of Supplementary Material). Even when the sensory input is noisy, the posterior means of the latent trajectories and the firing rates matched the ground truth well (Figure 3B–D). The PLNDE-inferred magnitudes of the noisy jumps induced by the pulses were similar to the true magnitudes (Pearson's $r = 0.714$), and the standard deviation of the posterior means was comparable to that of the true magnitudes (0.057 versus 0.059, respectively; Figure 3D).

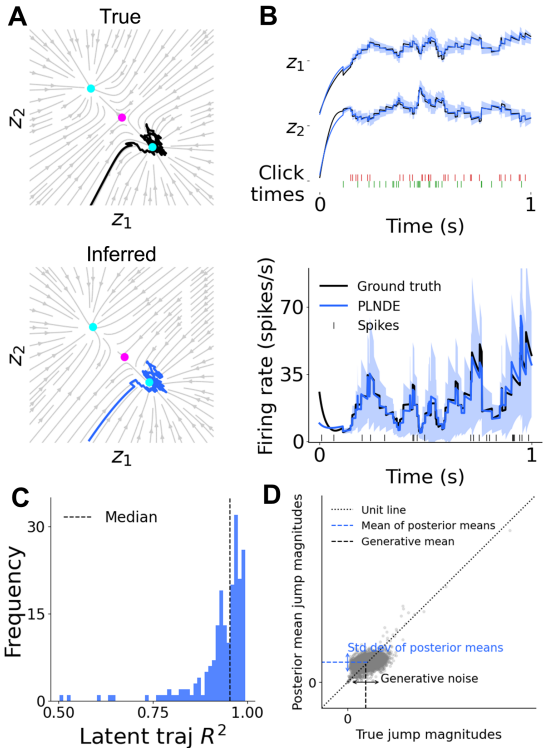

Figure 3. PLNDE can take discrete noisy sensory input stimuli. (A) Same as Figure 2A, but for mutual inhibition. (B) Black lines show a true example latent trajectory and firing rate while blue lines show the posterior means with ± 1.96 standard deviation tubes estimated from posterior samples. Red and green ticks indicate stimulus event times from the right and left sides, respectively. (C) Histogram of R^2 between the true test trajectories and the posterior mean (median=0.954). (D) Each grey circle represents the true magnitude of update in the latent state after a sensory event versus the posterior mean magnitude of update. The black arrow indicates ± 1.96 standard deviations around the true mean jump magnitudes and the blue arrow indicates ± 1.96 standard deviations around mean of the PLNDE posterior mean jump magnitudes.

4.4. Dynamics of the Medial Frontal Cortex During Perceptual Decision-Making

We fit PLNDE to a single session of recordings from the secondary motor cortex (M2), cingulate cortex (Cg1), prelimbic cortex (PrL) and medial orbital cortex (MO) in the mFC of a rat performing the decision-making task in Section 2.5 (see Figure 1B for this task; see Figure 4A for location of the probe). Only spikes that occurred during the stimulus period were used and the spike times were aligned to the onset of the stimulus. Spike times and click times from the total of 166 trials were used for fitting our model, and the model assumed that a pulse on either the left or the right side induces a jump in the latent dynamics exactly at the time when the pulse occurs. The configuration of the model was set to be the same as that in Section 4.3 (see Section B of Supplementary Material for details). We find that Cg1 contains the most information about choice and evidence compared to the other areas in the mFC, while PrL contains the least information about choice and evidence (Figure 4B–C). This was confirmed qualitatively (Figure 4B) and by performing linear regression to predict the difference in the total numbers of right and left clicks in each correct trial based on the final state of the latent trajectory (Figure 4C, top). The

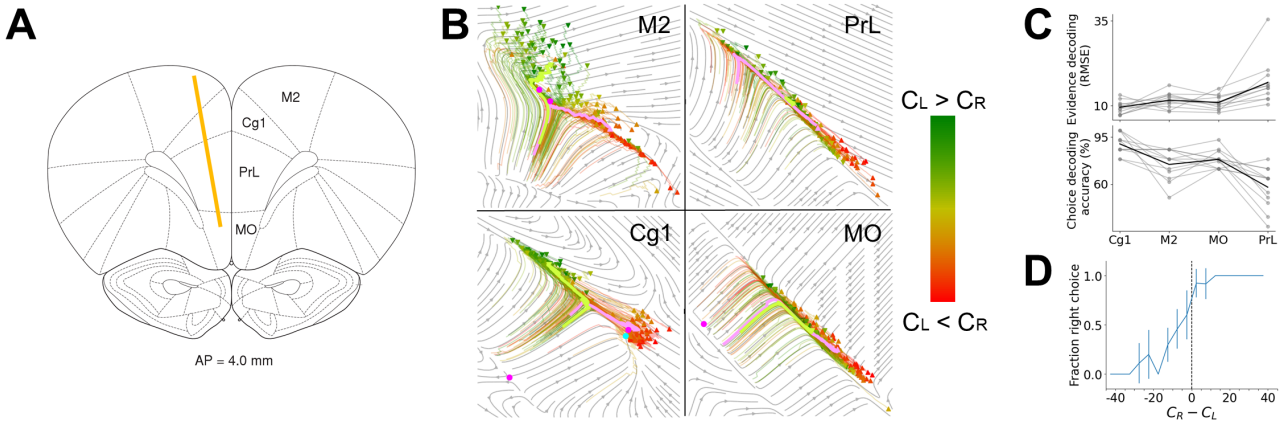

Figure 4. (A) Location of the Neuropixels probe in mFC. In total 497 neurons were recorded during the session with 101 M2 neurons, 99 Cg1 neurons, 249 PrL neurons, and 47 MO neurons. (B) Phase portraits of each area in mFC. Each trajectory is a single correct trial, color-coded such that green represents a trial with more clicks on the left side and red represents a trial with more clicks on the right side. Where each trajectory ends is marked with a triangle, oriented differently according to the rat's choice. Light green line is the trajectory that ends closest to the mean of the correct left choice trajectory final states, indicated by the light green triangle. Pink line is the trajectory that ends closest to the mean of the correct right choice trajectory final states, indicated by the pink triangle. Cyan and magenta circles indicate stable and unstable fixed points found by our model, respectively. (C) 12-fold cross-validated linear regressions for evidence decoding and linear SVMs for choice decoding. Each gray line indicates 1-fold. Black line is the mean. (D) Fraction of right choices is computed as a function of differences in the number of right and left clicks. Error bars indicate 95% binomial proportion confidence intervals.

Table 1. Summary of properties of latent variable models for spiking neural data

MODEL	PHASE PORTRAIT	INPUT STIMULI	SCALABLE[3]	CONTINUOUS-TIME	ASSUMPTIONS ON LATENTS
PLDS	√	√	√		LINEAR, STOCHASTIC
vLGP			√	√	NO DYNAMICS, STOCHASTIC
GP-SDE	√			√	POSTERIOR IS LTV[4], STOCHASTIC
LFADS[5]		√	√		NONLINEAR, DETERMINISTIC
PLNDE	√	√	√	√	NONLINEAR, STOCHASTIC[6]

decoding RMSE of PrL was significantly higher than all other areas ($p < 0.005$; two-sided Mann-Whitney U tests). We also trained a regularized linear support vector machine (SVM) to predict the rat's choice on each trial based on the final state of the latent trajectory (Figure 4C, bottom). Choice decoding accuracy of PrL was significantly lower than all other areas ($p < 0.01$; two-sided Mann-Whitney U tests), while decoding accuracy of Cg1 was significantly higher than all other areas ($p < 0.005$; two-sided Mann-Whitney U tests).

This result is consistent with our other work which showed that the dorsomedial frontal cortex (dmFC; area covering M2 and Cg1) contains more information about perceptual evidence than the medial prefrontal cortex (mPFC; area covering PrL and MO) (Luo et al., 2021).

Critically, PLNDE captures those results while also providing interpretable phase portraits of the latent space dynamics (Figure 4B). These portraits, including that of Cg1, which contains the most information about perceptual evidence and choice, suggest that at finite timescales per trial, a point that is stable to infinite time is not necessary to implement the animal's choice, in contrast to the view taken in Wong & Wang 2006.

5. Discussion

We introduced a general, interpretable framework that can be practically applied to neural population activities in a variety of tasks and species. In contrast to Gaussian process stochastic differential equation (Duncker et al. 2019; GP-SDE) or variational latent Gaussian process model (Zhao

[3]Check-marked if the model can deal with $L \gg 3$.
[4]Linear time-varying
[5]Can obtain fixed points and phase portraits in $\mathbb{R}^H$, but not $\mathbb{R}^L$.
[6]PLNDE latents are stochastic only at finite time points.

& Park 2017; vLGP), our model can take sensory inputs that are potentially noisy and discrete. Also, in contrast to LFADS or PLDS, our model can infer the nonlinear phase portraits, as demonstrated in synthetic neural population activities (Table 1). We believe that being able to plot the nonlinear phase portraits and being able to take sensory inputs are important because these will allow direct comparisons between phase portraits inferred from real neurophysiological data (e.g., Figure 4B) and traditional model phase portraits (e.g., Figure 5 in Wong & Wang 2006).

However, one should keep in mind that PLNDE does not provide a measure of confidence in its estimate of the phase portrait, and must be cautious when interpreting phase portrait far from the traversed trajectories. While PLNDE is a general framework, the phase portraits inferred from PLNDE may depend on specific model assumptions, including whether the dynamics are stationary or non-stationary, and how the sensory stimuli are represented in neural dynamics (Levi & Huk, 2020)—any interpretation we make from the inferred portraits is bound to reflect those assumptions. Therefore, before applying the model to the brain area of interest, one must consider the area's functional roles, and what form of sensory representation it gets.

We demonstrated in synthetic datasets that our model outperforms other models especially when there are fewer numbers of trials and spike counts. While LFADS has previously been applied mainly to data from monkeys, our approach may be advantageous for rodent data, as firing rates are often lower in rodents than monkeys for the analogous brain areas and in similar behaviors (see, for example, the rat posterior parietal cortex (PPC) mean firing rates in Hanks et al. 2015, and the monkey PPC mean firing rates in Roitman & Shadlen 2002). The number of trials available for analyses is often not large in a typical experimental session. LFADS proposes training with data from multiple sessions to increase inference power, and while this is also possible with our framework, our model can work session-by-session without overfitting even when the number of trials in each session is not large. One reason why our model performed well with small numbers of trials may be that the number of parameters of our model required to perform competitively against other models was relatively small (Section C of Supplementary Material). A large chunk of the parameters in LFADS goes to its recognition model. If we do have a sufficiently large number of trials, we may also consider amortizing posterior inference of our model by having, for example, an RNN encoder like (6), ODE-RNN encoder (Rubanova et al., 2019) or a NeuralCDE encoder (Kidger et al., 2020).

The network architecture for PLNDE was set to be the same throughout this paper. While this architecture gives reasonably good fits to our datasets, this may not be the best set of PLNDE hyperparameters. Similarly, all comparisons between PLNDE (with the architecture specified in Section B of Supplementary Material) and other models were carried out using the default hyperparameters given in their respective original repositories. Indeed, standardized routines that optimize the hyperparameters of these models will improve their performances in recovering the latent dynamics, especially in the regime where there are fewer numbers of trials and the models are more prone to overfitting. We suspect that hyperparameter optimization will be the most effective for LFADS as the number of parameters for LFADS is the largest and more than 100 times the numbers of parameters of the other models we compared with (Section C of Supplementary Material). Recent results suggest that the performance of LFADS increases with coordinated dropout (Keshtkaran & Pandarinath, 2019) and hyperparameter optimization (Keshtkaran et al., 2021), especially when there are small numbers of trials. Continued efforts to develop flexible and interpretable models of neural dynamics, and efforts to develop methods to discourage overfitting in these models so that they work out-of-the-box, will be valuable for systems neuroscience.

Acknowledgements

We thank Chethan Pandarinath, David Schwab, Lea Duncker, Yuan Zhao and the anonymous reviewers for their suggestions and comments. We also thank Abby Russo for letting us use her monkey illustration in Figure 1, and Chris Rackauckas for help with SciML. Rat experiments in this work were approved by the Princeton University Institutional Animal Care and Use Committee and were carried out in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. This work was supported by grants from NIH R01MH108358, the Simons Collaboration on the Global Brain (SCGB AWD543027), a U19 NIH-NINDS BRAIN Initiative Award (5U19NS104648), and by the Howard Hughes Medical Institute.

References

Brody, C. D., Romo, R., and Kepecs, A. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology*, 13(2):204–211, 2003.

Brunton, B. W., Botvinick, M. M., and Brody, C. D. Rats and humans can optimally accumulate evidence for decision-making. *Science*, 340(6128):95–98, 2013.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

Chaudhuri, R. and Fiete, I. Computational principles of memory. *Nature Neuroscience*, 19:394–403, 2016.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural Ordinary Differential Equations. *Advances in Neural Information Processing Systems*, 31: 6571–6583, 2018.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using rnn encoder-decoder for statistical machine translation, 2014.

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. Neural population dynamics during reaching. *Nature*, 487:51–56, 2012.

Corner, S., Sandu, C., and Sandu, A. Adjoint sensitivity analysis of hybrid multibody dynamical systems, 2018.

DePasquale, B. D., Pillow, J. W., and Brody, C. D. Accumulated evidence inferred from neural activity accurately predicts behavioral choice. In *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2019.

Duan, C. A., Pagan, M., Piet, A. T., Kopec, C. D., Akrami, A., Riordan, A. J., Erlich, J. C., and Brody, C. D. Collicular circuits for flexible sensorimotor routing. *Nature Neuroscience*, 2021.

Duncker, L., Bohner, G., Boussard, J., and Sahani, M. Learning interpretable continuous-time models of latent stochastic dynamical systems. *Proceedings of the 36th International Conference on Machine Learning*, 97:1726–1734, 2019.

E, W. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 5 (1), 2017.

Fetz, E. Are movement parameters recognizably coded in the activity of single neurons? *Behavioral and Brain Sciences*, 15(4):679–690, 1992.

Flint, R. D., Tate, M. C., Li, K., Templer, J. W., Rosenow, J. M., Pandarinath, C., and Slutzky, M. W. The representation of finger movement and force in human motor and premotor cortices. *eNeuro*, 7(4), 2020.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., and Brody, C. D. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, 520:220–223, 2015.

Harvey, C. D., Coen, P., and Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature*, 484:62–68, 2012.

Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.

Jia, J. and Benson, A. R. Neural jump stochastic differential equations. *Advances in Neural Information Processing Systems*, 2019.

Kaufman, M. T., Churchland, M. M., Ryu, S. I., and Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. *Nature Neuroscience*, 17(3): 440–448, 2014.

Keshtkaran, M. R. and Pandarinath, C. Enabling hyperparameter optimization in sequential autoencoders for spiking neural data. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

Keshtkaran, M. R., Sedler, A. R., Chowdhury, R. H., Tandon, R., Basrai, D., Nguyen, S. L., Sohn, H., Jazayeri, M., Miller, L. E., and Pandarinath, C. A large-scale neural network training framework for generalized estimation of single-trial population dynamics. *bioRxiv*, 2021.

Kidger, P., Morrill, J., Foster, J., and Lyons, T. Neural Controlled Differential Equations for Irregular Time Series. *Advances in Neural Information Processing Systems*, 2020.

Levi, A. J. and Huk, A. C. Interpreting temporal dynamics during sensory decision-making. *Current Opinion in Physiology*, 16:27–32, 2020.

Luo, T. Z., Hanks, T. D., Gupta, G., Bondy, A. G., and Brody, C. D. Dorsomedial frontal cortex participates in both evidence accumulation and trial history-based updating during perceptual decision-making. In *Computational and Systems Neuroscience (Cosyne) Abstracts*, 2021.

Machens, C. K., Romo, R., and Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005.

Macke, J. H., Buesing, L., Cunningham, J. P., Yu, B. M., Shenoy, K. V., and Sahani, M. Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems*, 24:1350–1358, 2011.

Mante, V., Sussillo, D., Shenoy, K. V., and Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, 2013.

Nieh, E. H., Schottdorf, M., Freeman, N. W., Low, R. J., Lewallen, S., Koay, S. A., Pinto, L., Gauthier, J. L., Brody, C. D., and Tank, D. W. Geometry of abstract learned knowledge in the hippocampus. *Nature*, 2021.

Pandarinath, C., O'Shea, D. J., Collins, J., et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, 15:805–815, 2018.

Piet, A. T., Erlich, J. C., Kopec, C. D., and Brody, C. D. Rat prefrontal cortex inactivations during decision making are explained by bistable attractor dynamics. *Neural Computation*, 29:2861–2886, 2017.

Piet, A. T., El Hady, A., and Brody, C. D. Rats adopt the optimal timescale for evidence integration in a dynamic environment. *Nature Communications*, 9, 2018.

Pinto, L., Koay, S. A., Engelhard, B., Yoon, A. M., Deverett, B., Thiberge, S. Y., Witten, I. B., Tank, D. W., and Brody, C. D. An accumulation-of-evidence task using visual pulses for mice navigating in virtual reality. *Frontiers in Behavioral Neuroscience*, 12:36, 2018.

Pontryagin, L., Boltyanskii, V., et al. (eds.). *The Mathematical Theory of Optimal Processes*. Interscience Publishers John Wiley & Sons, Inc., New York-London, 1962.

Rackauckas, C., Ma, Y., Dixit, V., Guo, X., Innes, M., Revels, J., Nyberg, J., and Ivaturi, V. A comparison of automatic differentiation and continuous sensitivity analysis for derivatives of differential equation solutions, 2018.

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., and Ramadhan, A. Universal differential equations for scientific machine learning. *arXiv preprint arXiv:2001.04385*, 2020.

Roitman, J. D. and Shadlen, M. N. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience*, 22(21):9475–9489, 2002.

Rubanova, Y., Chen, R. T. Q., and Duvenaud, D. K. Latent Ordinary Differential Equations for Irregularly-Sampled Time Series. *Advances in Neural Information Processing Systems*, 32:5320–5330, 2019.

Russo, A. A., Khajeh, R., Bittner, S. R., Perkins, S. M., Cunningham, J. P., Abbott, L. F., and Churchland, M. M. Neural trajectories in the supplementary motor area and motor cortex exhibit distinct geometries, compatible with different classes of computation. *Neuron*, 107(4):745–758, 2020.

Scott, B. B., Constantinople, C. M., Erlich, J. C., Tank, D. W., and Brody, C. D. Sources of noise during accumulation of evidence in unrestrained and voluntarily head-restrained rats. *eLife*, 4, 2015.

Serban, R. and Hindmarsh, A. C. CVODES: The sensitivity-enabled ode solver in sundials. In *Proceedings of the ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 6: 5th International Conference on Multibody Systems, Nonlinear Dynamics, and Control, Parts A, B, and C, pp. 257–269, 2005.

Seung, H. S. How the brain keeps the eyes still. *Proceedings of the National Academy of Sciences*, 93(23): 13339–13344, 1996.

Suresh, A. K., Goodman, J. M., Okorokova, E. V., Kaufman, M., Hatsopoulos, N. G., and Bensmaia, S. J. Neural population dynamics in motor cortex are different for reach and grasp. *eLife*, 9, 2020.

Sussillo, D. and Barack, O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Computation*, 25(3):626–649, 2013.

Sussillo, D., Jozefowicz, R., Abbott, L. F., and Pandarinath, C. Lfads - latent factor analysis via dynamical systems, 2016.

Vyas, S., Golub, M. D., Sussillo, D., and Shenoy, K. V. Computation through neural population dynamics annual review of neuroscience. *Annual Review of Neuroscience*, 43:249–275, 2020.

Wong, K. F. and Wang, X. J. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience*, 26(4):1314–1328, 2006.

Zhao, Y. and Park, I. M. Variational latent gaussian process for recovering single-trial dynamics from population spike trains. *Neural Computation*, 29(5):1293–1316, 2017.